Qeios

Peer Review

Review of: "Free Will: Reality and Perception"

Pedram Heydari[1]

1. Independent researcher

I really enjoyed reading this paper. It is dense and concisely written while being faithful to the nuances of the free-will debate, with memorable sentences such as “Once they wonder if or how they might get themselves to adopt a belief they don’t yet have, they stare into the abyss of infinite regress,” and “Tellingly, James left this be an experiment of thought instead of behavior.”

I have a few comments and clarification questions.

1. The null findings on moral behavior following reduced belief in free will are intriguing, but it is not clear that they are conclusive about longer-run effects. Most of the cited evidence measures behavior immediately after brief belief manipulations, and it seems plausible that behavior is less elastic than self-reported belief (especially for outcomes tied to habits, identity, and social reinforcement), so meaningful behavioral adjustment may require repeated exposure or longer time horizons. As a result, it remains unclear whether the short-run null effect would persist in the medium or long run, or whether delayed effects would emerge once beliefs have had time to shape norms, self-control strategies, or self-concept.
2. The discussion on overconfidence may benefit from citing the formal literature on instrumental belief distortions. For example, Bénabou & Tirole (2002, *Self-Confidence and Personal Motivation*) model a time-inconsistent but rational agent in which self-confidence influences effort decisions. In their framework, a forward-looking self has incentives to manage beliefs (e.g., by selectively attending to or interpreting discouraging signals) in order to sustain confidence and thereby motivate the short-run self’s exertion. As a result, overconfidence can arise endogenously in equilibrium.

doi.org/10.32388/44CASO

3. I was a little confused about the treatment of pragmatism and the scientific view in the section "Neither compatibilism nor pragmatism can rescue the freedom of the will." I elaborate on the confusion below:

   - The text alternates between "pragmatism as the payoff of believing" (a Belief → Good Behavior claim, naturally evaluated by P(g|B)) and "pragmatism as inferring truth from payoff" (a Good Behavior→Truth claim, evaluated by P(T|g)), without clearly distinguishing belief from ontological truth.

   - In the motivating slogans (Scientific view: "if a belief is true, then it is useful" and pragmatist view: "if a belief is useful, then it is true"), "true" is naturally read as a property of the proposition (e.g., whether free will actually exists). But in Figure 1 and the surrounding text ("half of the cases hold free will to be True, p(T)=0.5"), "True/False" reads more like a property of agents, i.e., whether a person endorses or holds free will to be true. If T is doxastic endorsement, then the conditionals being compared (P(g|T) and P(T|g)) concern behavior given belief and belief given behavior, rather than "reality ↔ usefulness," and the philosophical comparison should be framed accordingly. A quick fix would be to add one explicit sentence defining what T denotes in the tables (ontological truth vs. belief/endorsement), and then either (i) relabel T as "belief/endorsement" throughout and present the section as a social-psychological or rhetorical analysis, or (ii) introduce a separate variable for ontological truth and clarify that the tables are a toy hypothesis/evidence model rather than an empirical cross-tab of believers versus non-believers.

4. A key point that would benefit from clarification is the paper's suggestion that "pragmatists ... might seek to capitalize on this positive association by adding good behaviors," given the prior premise that people cannot change their beliefs "at will." As written, it is not clear what mechanism is supposed to generate more good behavior while the distribution of beliefs remains fixed (e.g., whether the author has in mind institutional incentives, social norms, policy interventions, selection effects, or something else). This matters because the right panel implements a very specific kind of "social improvement" (an increase in good cases applied across both the "True" and "False" columns). The subsequent conclusion that "a simple increase in good behavior is enough" may not hold for other plausible patterns of improvement—for instance, increases that primarily affect one belief group, or improvements that change the bad margin rather than adding good cases. It would be helpful to (i) state explicitly what class of interventions or base-rate shifts the right panel is intended to represent, and (ii) either provide a brief robustness argument (or examples) showing when the claimed ordering between the two views (scientific and pragmatist) persists and

when it does not. This would both sharpen the conceptual interpretation and appropriately scope the final claim.

Overall, it was a pleasure reading this stimulating and thought-provoking paper, and I hope these comments help further clarify and strengthen its arguments.

Declarations

Potential competing interests: No potential competing interests to declare.